Jury Gender Composition and Student Performance in Non-Blind Testing

Juan Muñoz-Morales *

September 26, 2024

Abstract

I employ candidate-level data from oral admission exams to assess how jury gender composition impacts student performance in non-blind testing. Students are assigned to evaluating committees quasi-randomly, ensuring an exogenous gender mix of jury members. A 10 percent increase in male committee members leads to a decrease of 0.02 standard deviations in scores, even after accounting for fixed effects. This decline is mainly driven by lower scores among male students evaluated by all-male committees. Evidence indicating that male jury members grade male candidates more harshly is provided. These results reveal noteworthy gender-based influences on student performance in non-blind testing.

Keywords: Gender, jury, admission exams, student performance.

JEL classifications: I23; J16; J71.

*IESEG School of Management, Univ. Lille, CNRS, UMR 9221- LEM-Lille Économie Management, F-59000 Lille, France (j.munoz@ieseg.fr). For helpful discussion and comments, I thank seminar participants at IFLAME and AEFP EdDev. I am especially grateful to Ghazala Azmat, Simone Moriconi, and Thomas Baudin.

1 Introduction

Screening mechanisms are widely employed in various contexts to address information asymmetries. For instance, recruiters consistently use tests and interviews to evaluate job-seekers, students, and grant applicants, whose qualifications they cannot fully observe. Although these assessments are useful screening techniques, they are susceptible to biases stemming from observable characteristics that do not accurately predict candidate quality.[1] Such biases can lead to inefficient outcomes, where the most qualified candidates are overlooked due to assessors incorrectly prioritizing characteristics such as gender or race, which do not indicate a candidate's suitability (Mechtenberg, 2009).

College admission exams are a prominent example of a screening mechanism that addresses the asymmetric information between colleges and prospective students. Recruiters do not fully observe applicants' skill levels; thus, admission exams serve as a strategy to accurately assess the skills of prospective students. These tests can be conducted either blindly (e.g., written exams where assessors cannot see candidates' characteristics) or non-blindly (e.g., oral exams where assessors can observe visible characteristics). However, non-blind testing can introduce potential discrimination based on assessors' biases.

This paper investigates whether non-blind testing in oral admission exams is influenced by recruiters' gender biases. Oral examinations are a specific form of assessment where irrelevant characteristics of the applicant are visible to the recruiter, allowing for potential biased evaluations based on preconceptions. For example, characteristics such as race and gender are evident during an oral examination and do not necessarily predict the candidate's quality. In fact, gender stereotypes in testing have been recognized as a significant source of bias (Carlana, 2019), and this could particularly affect oral admission exams. Gender biased assessments could have strong longer-term effects increasing gender inequality in access to elite institutions and further labor market outcomes.

I employ detailed data from oral admission exams at a post-secondary institution in France

[1] Several studies have identified biased assessments in multiple contexts. See, for instance: Anwar et al. (2012); Goldin and Rouse (2000); Broder (1993); Blank (1991); Card et al. (2019); Paola and Scoppa (2015); Bagues and Esteve-Volart (2010); and Bagues et al. (2017).

to explore the existence of gender biases in non-blind assessments. Student admission is contingent upon an oral examination conducted by committees that are quasi-randomly assigned, combining in-house (i.e., faculty members and administrative staff) and out-of-house (e.g., alumni, parents of students) jury members.[2] The gender composition of the committees is arranged in a manner that is uncorrelated with the evaluated students' quality, gender, or specific characteristics. Thus, I leverage the idiosyncratic variation in gender composition across committees to identify potential gender biases in oral assessments. I present strong empirical evidence suggesting that this gender composition is random and uncorrelated with baseline candidates' characteristics, including measures of cognitive and oral skills.

My findings reveal that a 10 percent increase in the proportion of male committee members results in a decrease of 0.02 standard deviations in oral test scores. This result remains robust when controlling for jury member fixed effects, indicating that assessors evaluate candidates differently based on the gender composition of their committee. Additionally, conditioning on candidates' baseline cognitive and oral skill levels does not alter the results, providing compelling evidence that the observed effects stem from gender biases among jury members. These findings hold across different committee sizes, alternative outcome measures, and can be replicated with other candidate samples.

Examining the heterogeneity of the results, I find that the negative impact is particularly pronounced for male students assigned to committees consisting entirely of male members. The results persist even when including jury member fixed effects, suggesting that male assessors penalize male candidates when no women are present in the committee. This aligns with previous studies that document biases against male candidates in public job recruitment (Bagues and Esteve-Volart, 2010) and in admission exams for post-secondary institutions (Breda and Ly, 2015; Breda and Hillion, 2016).

Two potential mechanisms may explain the main result. First, male assessors might grade male candidates more harshly in all-male committees. I find strong evidence supporting this hypothesis, as male assessors rate observably similar candidates lower when assigned to committees with a higher proportion of males. This situation is less clear for female assessors.

[2] Jury members deliberate to reach a final grade that determines a candidate's eligibility for admission.

Second, male candidates may feel intimidated and perform worse in committees with a higher male presence. While it is challenging to test this mechanism definitively, I provide indirect evidence indicating that the effects do not vary based on whether the oral test is conducted in person or online, or based on the candidates' oral skills. These two pieces of evidence suggest that male candidates do not alter their behavior, but the evidence is not enough to fully rule out this mechanism.

This paper contributes to the literature on assessment biases arising from factors that do not predict candidate quality. Such discrimination has been observed in various contexts, including criminal trials (Anwar et al., 2012) and labor market screenings (Goldin and Rouse, 2000; Bagues and Esteve-Volart, 2010). A specific area of this literature focuses on biases in academic careers, highlighting discrimination against certain minorities based on observable characteristics (Broder, 1993; Blank, 1991; Card et al., 2019; Paola and Scoppa, 2015; Bagues et al., 2017; Breda and Hillion, 2016). Other studies have found that individual assessments are negatively influenced by the quality of previously evaluated candidates (Radbruch and Schiprowski, 2024).

More specifically, this paper contributes to the literature quantifying discrimination in testing across various contexts. Race is a recurring characteristic that is subject to biases from graders (Hanna and Linden, 2012; Quinn, 2020; Burgess and Greaves, 2013; Botelho et al., 2015; Shi and Zhu, 2023; Sprietsma, 2013), as well as the alignment of characteristics between teachers and students (Gershenson et al., 2016; Lavy and Sand, 2018), and pre-existing beliefs about student behavior in the classroom (Ferman and Fontes, 2022).

Gender is another characteristic consistently identified as a source of bias. Mixed findings suggest the existence of negative biases towards either men or women. For example, Alan et al. (2018) find that girls are penalized when assessed by teachers with traditional gender views, whereas Jansson and Tyrefors (2022) show that female students' grades improve when graded anonymously.[3] Conversely, other studies suggest that male students are graded less favorably than equally skilled female students (Hinnerich et al., 2011; Terrier, 2020; Cornwell

[3]Some literature has focused on estimating gender discrimination in teacher evaluations rather than student assessments, showing a negative bias towards female teachers (Boring, 2017; Boring and Philippe, 2021; Lavy and Megalokonomou, ming).

et al., 2013; Falch and Naper, 2013; Lavy, 2008). Breda and Ly (2015) presents a closely related study that examines oral admissions exams in France, revealing grading biases that depend on whether the field is male- or female-dominated. The authors compare scores from blind and non-blind exams and find that female students are more likely to be admitted to male-dominated subjects. My contribution to this literature lies in demonstrating how assessors' evaluations change based on the gender composition of their assigned committee, including potential biases against male students evaluated by male assessors.

The results of this article also contribute to the literature on teacher biases, which encompass both grading and teaching. For instance, Carlana (2019) and Rakshit and Sahoo (2023) show that the gender gap in math increases when students are assigned to teachers with gender stereotypes. Additionally, Hoffmann and Oreopoulos (2009) and Dee (2007) find that having a teacher of the same gender improves student achievement, while Papageorge et al. (2020) suggests that teachers' expectations significantly impact students' college completion rates. This paper adds to this body of literature by highlighting how interactions during a brief interview can influence how assessors grade students based on observable characteristics.

Finally, this article relates to the literature on decision-making by committees. Theoretically, this topic has been explored by analyzing the optimal design of voting rules and the roles of communication and ex-ante commitments.[4] Empirically, Goeree and Yariv (2011) and Iaryczower et al. (2018) investigate committee decisions with and without deliberation, Chan (2021) analyzes the impact of seniority on decisions made by physician committees, and Radbruch and Schiprowski (2023) studies the influence of women's recommendations on final committee decisions. Two key findings from this literature are relevant: first, deliberation significantly affects the decisions made by committees; second, the assessments of certain committee members (such as more senior members and men) carry disproportionate weight in final decisions. My results contribute to this discourse by illustrating how committee decisions may also vary based on gender composition and how assessments by jury members might change depending on the proportion of males in the committee.

[4]See, for instance: (Li and Suen, 2009).

2 Setting and Data

2.1 Setting

This analysis focuses on students who applied in 2023 for admission to a business major program at a post-secondary institution in France. The admission process comprised two stages. First, students took written baseline exams in numeracy, literacy, and foreign language. These exams consisted mainly of multiple-choice questions that were graded blindly. Second, top performers from the written exams were selected to take oral admission examinations, which consisted of two parts: (1) a *foreign language* component and (2) a *core* component.[5] Students could take either component at two different sites or online, and scheduling was available for both morning and afternoon sessions. Final eligibility was determined by a weighted average of the written baseline exams, the foreign language component, and the core oral component.

The *foreign language* oral component involved an interview with an expert who assessed the candidate's proficiency in the language. Candidates were evaluated orally by a single assessor, who assigned a score to each student. This exam was conducted independently from the *core* component and could even be scheduled on a different date; however, most students opted to take both exams on the same day.

The *core* component of the oral exam assessed candidates' (1) general knowledge and (2) motivation to enter the school. Each candidate was assigned to a committee that determined eligibility based on these two criteria. At the end of the test, each candidate received a single grade from the committee.

Committees: Students' scores in the *core* component were determined by committees composed of two or three jury members. Jury members were either school personnel (i.e., in-house members) or individuals affiliated with the school (i.e., out-of-house members). In-house members included current students in advanced years, administrative staff, or faculty. Out-of-house jury members comprised alumni, parents of current students, or professionals from organizations connected to the school.

[5] Students who scored above 60 percent on the written exam were eligible to participate in the oral exams.

Prior to the oral exams, the school designated specific days for the examinations. A total of 10 days, with morning and afternoon sessions, were offered. Jury members signed up to participate in the oral examinations on these predetermined dates. In-house jury members were typically required to attend three days of oral exams, equating to six sessions, while out-of-house jury members were usually assigned two to three sessions.

On the day of the exam, jury members were assigned to committees. Any jury member knew about its committee until the beginning of the evaluating session. Each committee evaluated four candidates over a four-hour period, dedicating 40 minutes to the exam and 20 minutes to deliberation.[6] During the 40 minutes of examination, 20 minutes were allocated to general knowledge assessment, while the remainder focused on assessing motivation to join the school. At the conclusion of each four-hour session, each committee reported the final grade for each candidate, after which the committee was dissolved.

Jury Assignment: Jury members were assigned to different committees for every session attended in an *as-good-as-random* manner. Three criteria guided the assignment of jury members to committees. First, jury members could not be assigned to the same peers more than once. Second, each committee was required to include both in-house and out-of-house jury members, ensuring at least one out-of-house member per committee. Third, efforts were made to achieve gender parity, though this was often not realized. No-shows, the requirement for at least one out-of-house member per committee, and insufficient jury members from a given gender complicated the goal of maintaining gender parity in all committees. Additionally, committees consisting of three members (assigned also quasi-randomly) could not, by definition, comply with the gender parity rule. Consequently, the gender composition of committees was assigned quasi-randomly among those where gender parity was not achieved. I provide empirical evidence of the quasi-random nature of jury assignment in section 4.1.

This assignment mechanism guarantees two key points relevant to the objective of this paper. First, the gender composition of the committee was not correlated with the quality or

[6]Some committees evaluated fewer candidates when students did not show up or due to insufficient candidates assigned.

gender of the candidate. Second, jury members were assigned to multiple committees with plausibly different gender compositions, creating within-jury member variation in the gender composition of the committees to which they were assigned.

2.2 Data

The data consist of a combination of oral admission test scores, baseline written scores, and characteristics of the candidates, juries, and committees. For every candidate, I observe their baseline written test score measures, the oral *core* and *foreign language* test scores, the committee to which they were assigned, and basic demographic characteristics.[7] A total of 2,611 students were evaluated by 694 juries over nine days, across 862 committees.[8]

A description of the sample used is presented in Appendix Table 1. Panel A describes the committees (at the student level), panel B the sample of students, and Panel C the sample of jury members. The share of males in each evaluating committee varies from zero to one, with a significant concentration around 0.5, indicating a substantial number of committees with gender parity. Approximately 14 percent of the committees had three jury members, and about 61 percent of the juries were in-house. Furthermore, the sample of students is predominantly male, with only 39 percent being female candidates.

Appendix Table 2 further describes the 862 committees and their composition. Columns (1) to (6) present descriptive statistics for all the committees, while columns (7) to (11) split the sample by the share of males. Importantly, about 39 percent (334 out of 862) of committees did not achieve gender parity; 10 percent were fully composed of males, and 14 percent were composed entirely of females. These values suggest that there is sufficient variation in gender composition across committees.

3 Empirical Strategy

I exploit the idiosyncratic variation in jury assignment to estimate a cross-sectional linear model that quantifies the effect of jury gender composition on students' scores. Formally, the model is expressed as:

[7] Baseline scores correspond to blind scores in the written exam (numeracy, literacy, and foreign language).
[8] I exclude 183 observations for which it was not possible to identify all the jury members.

$$y_{iJ} = \alpha + \beta S_J + \delta X_i + \gamma T_i + \varepsilon_{iJ}, \tag{1}$$

where y_{ij} corresponds to the score of student i assigned to committee J, S_J is the share of males in the committee, X_i is a matrix of student-level characteristics, T_i includes baseline student test scores, and ε_{iJ} is an idiosyncratic error clustered at the committee level. Importantly, I observe student scores only at the candidate level (i.e., the score given by the committee) and not the test score given by each individual jury member.

Many characteristics of the jury members, aside from gender, can create biases and affect students' scores. Therefore, I stack student-by-jury-member observations and estimate an alternative model that conditions on jury member fixed effects, μ_j.[9] Formally, this second model is represented as:

$$y_{iJ} = \alpha + \beta S_J + \delta X_i + \gamma T_i + \mu_j + \varepsilon_{ij}. \tag{2}$$

The specification in Equation 2 compares students assigned to the same jury member, j, but in different committees, J, which may vary in gender composition. Recall that jury members are assigned to different committees to assess various candidates. Including the jury member fixed effect uses within-jury-member variation to control for any jury member-specific characteristic that might affect the assessments and compares the student scores given by the same juries assigned to different committees, which could vary in their gender composition.

4 Results

4.1 Validity of the Research Design

If jury members are assigned in an as-good-as-random fashion, then the share of males per committee is expected to be uncorrelated with candidates' characteristics, and committees with and without gender parity should be similar in terms of observable characteristics. A lack of significant point estimates implies that candidates are not assigned to specific committees in a particular way, providing strong evidence against selection bias in the assignment of jury

[9] The outcome of the estimation varies exclusively by student and not by student-by-jury member given that I do not observe the score given by each jury member.

members to candidates.

To test this, I estimate Equation 1 and 2 using candidates' characteristics as outcomes, and present the results in Figure 1a.[10] The share of males does not systematically correlate with any of the observed characteristics, suggesting that there is no selection bias in the assignment of candidates to juries. Specifically, this share does not correlate with baseline test score measures, providing strong evidence for the exogeneity of the measure. These results hold across specifications with and without jury member fixed effects.

Additionally, I compute regressions at the committee level using multiple outcomes for each committee on a dummy variable that takes the value of one if the committee had gender parity. The results are displayed in Figure 1b. There are no significant differences between committees with and without gender parity, further supporting the quasi-random nature of the jury assignments.

4.2 Effect of Gender Composition on Scores

Students' scores in the *core* component of the oral admission exam exhibit a negative correlation with the gender composition of the committees, particularly among male candidates. Figure 2 presents the raw standardized averages in the *core* oral admission exam based on the share of males in each committee, differentiated by candidates' gender. The test scores show a monotonically decreasing trend for male candidates, whereas the scores for female candidates do not display the same pattern. Notably, male candidates assigned to committees with a higher proportion of male jury members tend to perform worse than their peers in more gender-diverse committees.

I provide more robust evidence for this assertion in Table 1, where I present the results of estimating Equations 1 and 2, using scores from the oral *core* exam as the dependent variable and conditioning on baseline characteristics and test scores. The odd-numbered columns show results pooled across genders, while the even-numbered columns interact the share of males with candidates' gender.

[10] The full results are also presented in Appendix Tables 3 and 4.

A 10 percent increase in the share of males corresponds to an average decrease of approximately 0.02 standard deviations in student test scores, with the negative effects predominantly affecting male candidates.[11] These coefficients remain stable across the various specifications that control for either baseline characteristics or test scores. Furthermore, I present the results from Equation 2 using within-jury-member variation in Panel B, where the point estimates remain consistently stable, even when accounting for time-invariant characteristics of the jury members.

These negative effects of the jury's gender composition affect particularly male candidates assigned to committees composed entirely of males. Figure 3 displays point estimates from Equation 1, with the gender composition variable discretized and the sample separated by male and female candidates.[12] The negative point estimates are found exclusively among male students assigned to all-male committees, with no adverse effects noted among female candidates.

Overall, these results indicate that the share of males in admission committees negatively influences students' scores. This effect is primarily driven by male candidates assigned to committees composed entirely of males. Additionally, the magnitude of the effect remains unchanged when controlling by assessors' specific attributes captured by the jury member fixed effects.

4.3 Robustness of the Effect

Two Jury Members: The results previously discussed may be confounded by the size of the committee. If the observed effects are solely driven by committees with three members, this could confound the actual impact of gender composition on student performance. To address this, I present the main results restricted to committees with two members in Appendix Table 5. The point estimates remain very similar in magnitude, although some precision is lost—particularly in Panel A —-due to the reduced number of observations.

[11] I cannot reject the null hypothesis that the effect differs between male and female candidates when controlling for individual characteristics or test scores in columns (4) and (6). This may be due to a lack of statistical power in the estimation. Nevertheless, the point estimates across all specifications for male candidates are negative and significantly different from zero, while for female candidates, the estimates do not show the same significance.

[12] Additionally, I present estimates pooled across male and female candidates in Appendix Figure 1, and separately by gender, but with jury member fixed effects, in Appendix Figure 2.

Admission to Further Years: There is a possibility that the results are merely coincidental within this specific sample of candidates. To investigate this, I estimate Equation 1 using the scores from oral admission exams for the same program but for students in subsequent years. Candidates can also be admitted starting in their second or third year, and they must take an oral exam as part of the admission process. For this sample, baseline test scores are not available, as these students are admitted solely based on their oral examination results. Appendix Table 6 presents the results, which again reveal negative and robust point estimates similar in magnitude to those in Table 1. This suggests that even when focusing on an alternative sample of older students, a larger share of males in committees is still associated with lower scores in the oral admission exam.

Using the difference between Blind and Non-blind Scores:- Finally, I follow the approach of Breda and Ly (2015) to compute a measure of "examiners' gender bias," defined as the gap between oral and written test scores.[13] This measure is used as the outcome variable in Equations 1 and 2. It is important to interpret these results with caution, as this measure captures examiners' bias only if the written and oral skills of the students remain constant, which is a strong assumption. The results are presented in Appendix Table 7, where we see that correcting for this measure does not alter any of the findings.

5 Mechanisms

The proportion of males in each committee significantly affects student performance on the exam. Two potential mechanisms could drive this effect. First, male jury members might grade male candidates more harshly in all-male committees. Second, male candidates could feel intimidated and perform worse in the presence of more male jury members. While it is challenging to empirically test these mechanisms, I provide strong evidence supporting the first and inconclusive evidence regarding the second.

[13] The examiners' bias is calculated as the residuals from a regression of the *core* oral test score on the average of the blind test scores obtained in numeracy, literacy, and foreign language.

5.1 Do Male Juries Assess Differently by Male Candidates?

Male jury members may adopt a harsher grading approach when assigned to committees with a higher proportion of males. To investigate this, I estimate Equation 2 and differentiate the results based on the gender of the jury members. The findings are presented in Table 2. Recall that estimations including jury member fixed effects compare students assigned to the same jury member across different committees with varying gender compositions.

The results show negative and significant point estimates for male jury members, while point estimates for female jurors are insignificant. This suggests that male jury members assess male candidates more harshly when they are in committees with a larger share of males. Female jury members also appear to grade male candidates more harshly in similar conditions, as indicated by the negative point estimates in Panel B; however, these estimates are not statistically distinguishable from zero.[14]

5.2 Do Male Candidate Performance Decrease?

The main results could also stem from changes in male candidates' performance based on the share of males in the committee. If male candidates feel coerced by a higher presence of males, they might perform worse than in other scenarios. While direct testing of this phenomenon is not feasible with the current data, I explore two indirect tests that provide evidence against this hypothesis.

First, if male candidates are indeed coerced in the presence of more male jury members, we would expect differences in performance between male and female candidates when taking the oral exam in person versus online. However, this does not appear to be the case. I interact the share of males with whether the interview was conducted online and present the results in Appendix Figure 3, split by candidates' gender.[15] The effect is consistent for both male and female candidates, suggesting that the share of males in the committee reduces test scores uniformly, regardless of whether the exam was conducted in person or online.

[14]I cannot statistically reject the null hypothesis that the point estimates in columns (1), (3), and (5) differ between Panels A and B, despite the coefficients for male jury members being approximately twice as negative as those for female jury members. This may be due to a lack of statistical power.

[15]Appendix Table 8 includes the results of the estimation with interaction terms, along with heterogeneous effects based on additional baseline characteristics for completeness.

Second, if coercion is a factor, one might expect that male candidates with better oral skills would suffer less when exposed to more male jury members. However, this also does not seem to be the case. I test this hypothesis by using the test score in the oral foreign language exam as a proxy for oral skills and estimating the heterogeneous effects by interacting the share of males with this measure, and splitting again by candidates' gender. The results are shown in Appendix Figure 4, where I again find that the effect does not vary by test score level or gender.

Overall, these results indicate that neither female nor male candidates are differentially affected by the presence of male jury members, whether taking the exam online or based on their oral skills (as measured by the oral foreign language exam score). Although these two pieces of evidence do not directly demonstrate that male students alter their behavior when exposed to more male jury members, they provide valuable support against this phenomenon.

6 Conclusions

This paper examines how non-blind testing impacts student scores in oral admission exams for post-secondary education. Candidates in this setting apply for admission by taking baseline blind exams alongside oral, non-blind assessments. The oral exams are evaluated by committees with gender compositions that vary exogenously. By exploiting this idiosyncratic variation, I identify the effect of committee gender composition on candidates' test scores in the oral admission exam.

The results indicate that an increase in the proportion of male committee members leads to lower scores for candidates, particularly for male candidates assigned to committees composed entirely of male jury members. This finding remains robust when controlling for variation across jury members, suggesting that assessors evaluate candidates differently based on the gender composition of their committee. Additionally, I find evidence that male assessors grade male candidates more harshly in male-dominated committees. Conversely, there is no evidence suggesting that male candidates feel intimidated by a male-dominated jury, although the results regarding this aspect are rather inconclusive.

These findings underscore the significant impact that assessor biases can have when grading students in non-blind exams. Jury members may judge different candidates based on observable characteristics that do not accurately reflect their abilities. This can have long-lasting implications for students who are not selected for specific programs, potentially leading to enduring disadvantages. Investigating these long-term effects, particularly in terms of labor market outcomes, remains an open question.

References

Alan, S., Ertac, S., and Mumcu, I. (2018). Gender Stereotypes in the Classroom and Effects on Achievement. *The Review of Economics and Statistics*, 100(5):876–890.

Anwar, S., Bayer, P., and Hjalmarsson, R. (2012). The Impact of Jury Race in Criminal Trials. *The Quarterly Journal of Economics*, 127(2):1017–1055.

Bagues, M., Sylos-Labini, M., and Zinovyeva, N. (2017). Does the gender composition of scientific committees matter? *American Economic Review*, 107(4):1207–38.

Bagues, M. F. and Esteve-Volart, B. (2010). Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment. *The Review of Economic Studies*, 77(4):1301–1328.

Blank, R. M. (1991). The effects of double-blind versus single-blind reviewing: Experimental evidence from the american economic review. *The American Economic Review*, 81(5):1041–1067.

Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics*, 145:27–41.

Boring, A. and Philippe, A. (2021). Reducing discrimination in the field: Evidence from an awareness raising intervention targeting gender biases in student evaluations of teaching. *Journal of Public Economics*, 193:104323.

Botelho, F., Madeira, R. A., and Rangel, M. A. (2015). Racial discrimination in grading: Evidence from brazil. *American Economic Journal: Applied Economics*, 7(4):37–52.

Breda, T. and Hillion, M. (2016). Teaching accreditation exams reveal grading biases favor women in male-dominated disciplines in france. *Science*, 353(6298):474–478.

Breda, T. and Ly, S. T. (2015). Professors in core science fields are not always biased against women: Evidence from france. *American Economic Journal: Applied Economics*, 7(4):53–75.

Broder, I. E. (1993). Review of nsf economics proposals: Gender and institutional patterns. *American Economic Review*, 83(4):964–70.

Burgess, S. and Greaves, E. (2013). Test scores, subjective assessment, and stereotyping of ethnic minorities. *Journal of Labor Economics*, 31(3):535–576.

Card, D., DellaVigna, S., Funk, P., and Iriberri, N. (2019). Are Referees and Editors in Economics Gender Neutral?*. *The Quarterly Journal of Economics*, 135(1):269–327.

Carlana, M. (2019). Implicit Stereotypes: Evidence from Teachers' Gender Bias*. *The Quarterly Journal of Economics*, 134(3):1163–1224.

Chan, D. C. (2021). Influence and information in team decisions: Evidence from medical residency. *American Economic Journal: Economic Policy*, 13(1):106–37.

Cornwell, C., Mustard, D. B., and Parys, J. V. (2013). Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *Journal of Human Resources*, 48(1):236–264.

Dee, T. S. (2007). Teachers and the gender gaps in student achievement. *Journal of Human Resources*, XLII(3):528–554.

Falch, T. and Naper, L. R. (2013). Educational evaluation schemes and gender gaps in student achievement. *Economics of Education Review*, 36:12–25.

Ferman, B. and Fontes, L. F. (2022). Assessing knowledge or classroom behavior? evidence of teachers' grading bias. *Journal of Public Economics*, 216:104773.

Gershenson, S., Holt, S. B., and Papageorge, N. W. (2016). Who believes in me? the effect of student–teacher demographic match on teacher expectations. *Economics of Education Review*, 52:209–224.

Goeree, J. K. and Yariv, L. (2011). An experimental study of collective deliberation. *Econometrica*, 79(3):893–921.

Goldin, C. and Rouse, C. (2000). Orchestrating impartiality: The impact of "blind" auditions on female musicians. *American Economic Review*, 90(4):715–741.

Hanna, R. N. and Linden, L. L. (2012). Discrimination in grading. *American Economic Journal: Economic Policy*, 4(4):146–68.

Hinnerich, B. T., Höglin, E., and Johannesson, M. (2011). Are boys discriminated in swedish high schools? *Economics of Education Review*, 30(4):682–690.

Hoffmann, F. and Oreopoulos, P. (2009). A professor like me. *Journal of Human Resources*, 44(2):479–494.

Iaryczower, M., Shi, X., and Shum, M. (2018). Can words get in the way? the effect of deliberation in collective decision making. *Journal of Political Economy*, 126(2):688–734.

Jansson, J. and Tyrefors, B. (2022). Grading bias and the leaky pipeline in economics: Evidence from stockholm university. *Labour Economics*, 78:102212.

Lavy, V. (2008). Do gender stereotypes reduce girls' or boys' human capital outcomes? evidence from a natural experiment. *Journal of Public Economics*, 92(10):2083–2105.

Lavy, V. and Megalokonomou, R. (Forthcoming). The short- and the long-run impact of gender-biased teachers. *American Economic Journal: Economic Policy*.

Lavy, V. and Sand, E. (2018). On the origins of gender gaps in human capital: Short- and long-term consequences of teachers' biases. *Journal of Public Economics*, 167:263–279.

Li, H. and Suen, W. (2009). Viewpoint: Decision-making in committees. *The Canadian Journal of Economics / Revue canadienne d'Economique*, 42(2):359–392.

Mechtenberg, L. (2009). Cheap Talk in the Classroom: How Biased Grading at School Explains Gender Differences in Achievements, Career Choices and Wages. *The Review of Economic Studies*, 76(4):1431–1459.

Paola, M. D. and Scoppa, V. (2015). Gender discrimination and evaluators' gender: Evidence from italian academia. *Economica*, 82(325):162–188.

Papageorge, N. W., Gershenson, S., and Kang, K. M. (2020). Teacher Expectations Matter. *The Review of Economics and Statistics*, 102(2):234–251.

Quinn, D. M. (2020). Experimental evidence on teachers' racial bias in student evaluation: The role of grading scales. *Educational Evaluation and Policy Analysis*, 42(3):375–392.

Radbruch, J. and Schiprowski, A. (2023). Committee Deliberation and Gender Differences in Influence. CRC TR 224 Discussion Paper Series crctr224_2023_430, University of Bonn and University of Mannheim, Germany.

Radbruch, J. and Schiprowski, A. (2024). Interview Sequences and the Formation of Subjective Assessments. *The Review of Economic Studies*, page rdae039.

Rakshit, S. and Sahoo, S. (2023). Biased teachers and gender gap in learning outcomes: Evidence from india. *Journal of Development Economics*, 161:103041.

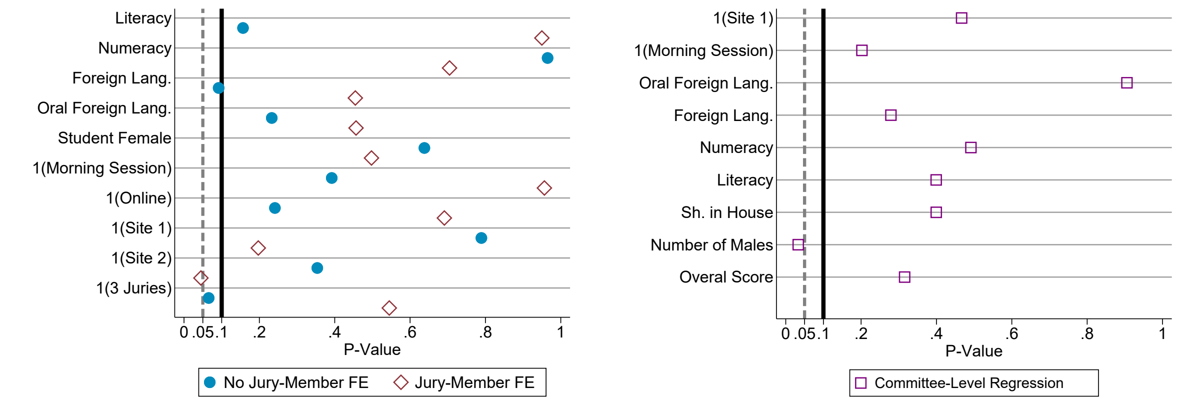
Shi, Y. and Zhu, M. (2023). "model minorities" in the classroom? positive evaluation bias towards asian students and its consequences. *Journal of Public Economics*, 220:104838.

Sprietsma, M. (2013). Discrimination in grading: experimental evidence from primary school teachers. *Empirical Economics*, 45(1):523–538.

Terrier, C. (2020). Boys lag behind: How teachers' gender biases affect student achievement. *Economics of Education Review*, 77:101981.

Figure 1: Balance Across Student and Committee Characteristics

(a) Student-Level (b) Committee-Level



Notes. This figure presents p-values of individual regressions on the outcomes described at the left of each panel. Panel 1a (N=2,611) uses the share of male jury members as independent variable. Blue dots represent p-values of estimations that do not include any covariate and are estimated at the student level. Red dots represent p-values of point estimates estimated at the student-jury member level and include jury fixed effects. Panel 1b (N=N = 862) presents the p-values of regressions at the committee level, where the covariate in the y-axis is regressed on a dummy that takes the value of one if the committee has gender parity. Standard errors are always clustered at the committee level.

Figure 2: Raw Scores by Gender and Committee Gender Composition

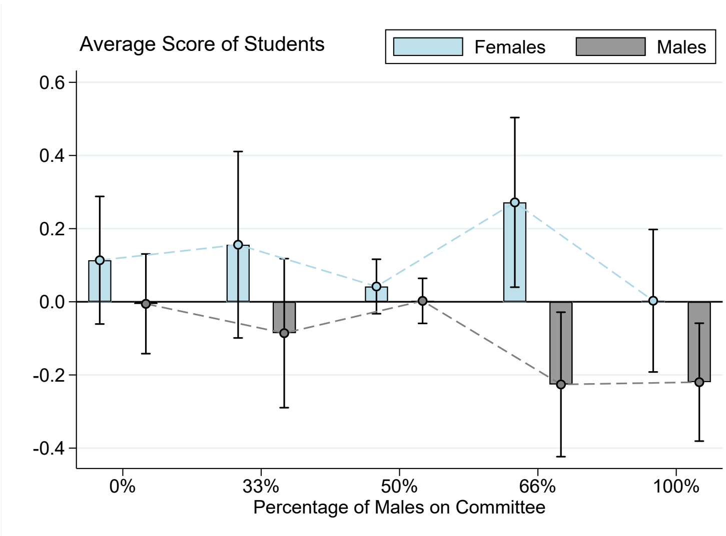
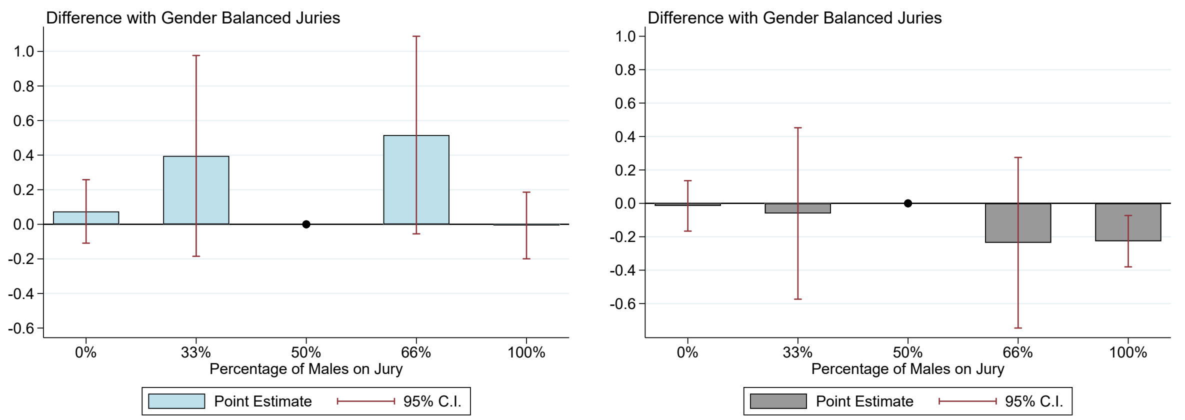


Figure 3: Differential Effects with Respect to Gender Balanced Committees by Candidates' Gender

(a) Female Candidates (b) Male Candidates



Notes. These figures present the point estimates of Equation 1 using the share of males discretized. Juries with 50 percent males are used as omitted category. Panel 3a conditions on female students, whereas Panel 3b on male students. Standard errors are clustered at committee level.

Table 1: Effects of Jury Gender Composition on Oral Test Scores

	(1)	(2)	(3)	(4)	(5)	(6)
A) No Jury Member Fixed-Effect						
Share of Males (β)	-0.161**	-0.242***	-0.147*	-0.193*	-0.159**	-0.207**
	(0.077)	(0.082)	(0.076)	(0.099)	(0.075)	(0.098)
Share of Males*1(female) (γ)		0.213***		0.119		0.129
		(0.071)		(0.157)		(0.154)
Women ($\beta + \gamma$)		-0.0288		-0.0746		-0.0774
p-value		0.745		0.536		0.516
Observations	2,611	2,611	2,611	2,611	2,610	2,610
B) Jury Member Fixed-Effect						
Share of Males (β)	-0.205**	-0.295***	-0.213***	-0.286***	-0.218***	-0.287***
	(0.080)	(0.086)	(0.078)	(0.106)	(0.080)	(0.107)
Share of Males*1(female) (γ)		0.254***		0.179		0.178
		(0.081)		(0.172)		(0.168)
Women ($\beta + \gamma$)		-0.0405		-0.107		-0.109
p-value		0.665		0.401		0.383
Observations	5,584	5,584	5,584	5,584	5,582	5,582
Candidate Controls			Yes	Yes	Yes	Yes
Baseline Test Scores					Yes	Yes

Note: This table presents the results of the estimation of Equation 1 (Panel A) and Equation 2 (Panel B) using the standardized oral exam test score as dependent variable. The share of males corresponds to the percentage of male jury members in each committee. Columns (1) and (2) have no controls, columns (3) and (4) include candidate-level controls, and columns (5) and (6) include candidate-level controls and test score measures. Candidate-level controls include gender, if the exam was performed in the morning, site fixed effects (online, site one, and site two), if the committee had three members, the share of jury members that are part of the school, and the share of jury members that are part of the school's faculty. Test score measures include the candidate's standardized result in the blind (reading, math, and English) and non-blind exams (oral English exam). Specifications in even columns fully interact the share of males and the controls with a dummy variable of whether or not the candidate is female. Standard errors are clustered at the committee level. *** p<0.01, ** p<0.05, * p<0.1.

Table 2: Effects of Jury Gender Composition By Gender of the Jury-Member

	(1)	(2)	(3)	(4)	(5)	(6)
A) Male Juries						
Share of Males (β)	-0.265**	-0.356***	-0.249**	-0.458***	-0.263**	-0.474***
	(0.115)	(0.119)	(0.112)	(0.152)	(0.114)	(0.156)
Share of Males*1(female) (γ)		0.264***		0.449*		0.471*
		(0.083)		(0.252)		(0.245)
Women ($\beta + \gamma$)		-0.0915		-0.00852		-0.00307
p-value		0.460		0.964		0.987
Observations	2,737	2,737	2,737	2,737	2,735	2,735
B) Female Juries						
Share of Males (β)	-0.147	-0.231*	-0.161	-0.154	-0.167	-0.144
	(0.119)	(0.125)	(0.118)	(0.157)	(0.119)	(0.157)
Share of Males*1(female) (γ)		0.232**		-0.009		-0.029
		(0.104)		(0.251)		(0.246)
Women ($\beta + \gamma$)		0.00141		-0.163		-0.173
p-value		0.992		0.387		0.356
Observations	2,847	2,847	2,847	2,847	2,847	2,847
Candidate Controls			Yes	Yes	Yes	Yes
Baseline Test Scores					Yes	Yes

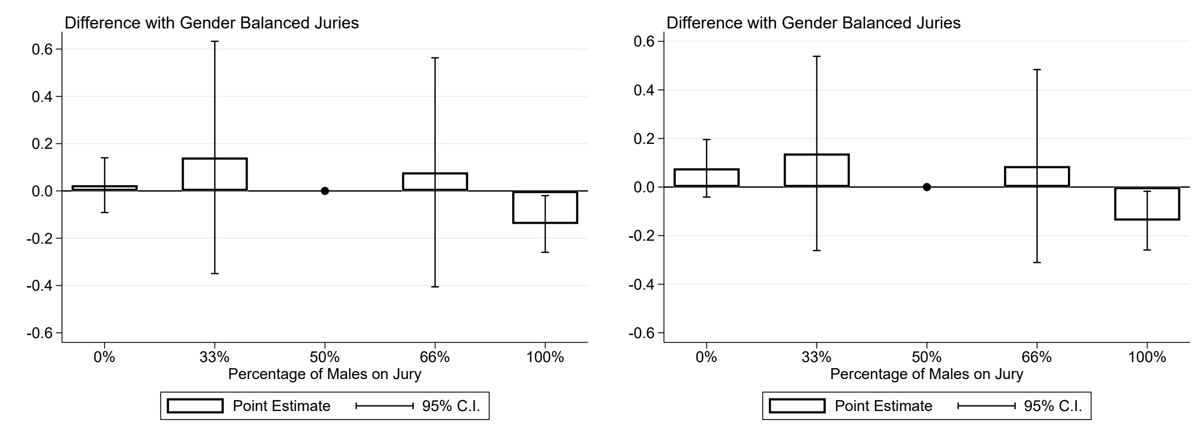
Note: This table presents the results of the estimation of Equation 2 using the standardized oral exam test score as dependent variable and restricting to male jury-members (Panel A) and female jury-members (Panel B), separately. The share of males corresponds to the percentage of male jury members in each committee. Columns (1) and (2) have no controls, columns (3) and (4) include candidate-level controls, and columns (5) and (6) include candidate-level controls and test score measures. Candidate-level controls include gender, if the exam was performed in the morning, site fixed effects (online, site one, and site two), if the committee had three members, the share of jury members that are part of the school, and the share of jury members that are part of the school's faculty. Test score measures include the candidate's standardized result in the blind (reading, math, and English) and non-blind exams (oral English exam). Specifications in even columns fully interact the share of males and the controls with a dummy variable of whether or not the candidate is female. Standard errors are clustered at the committee level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Appendix: Additional Figures and Tables

Appendix Figure 1: Differential Effects with Respect to Gender Balanced Committees

(a) No Jury Member Fixed Effect

(b) Jury Member Fixed Effect

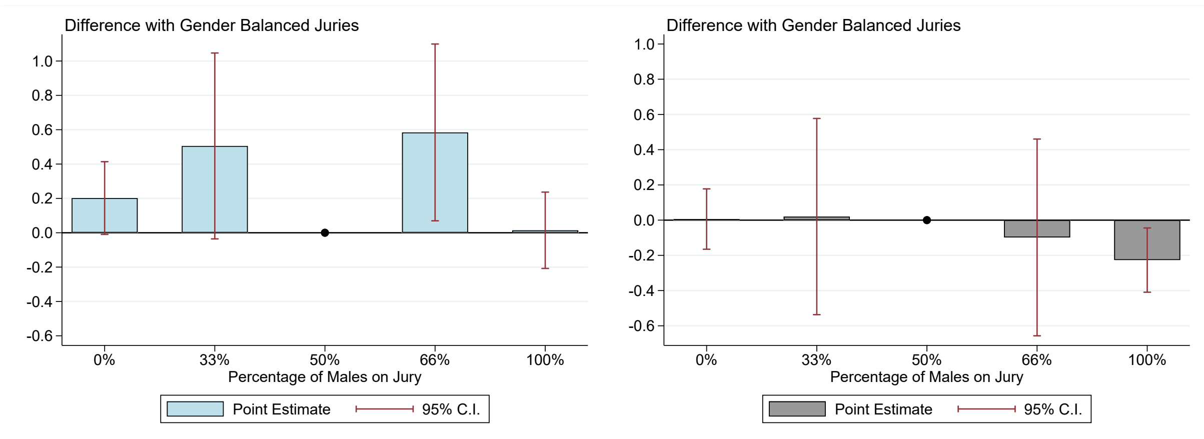


Notes. These figures present the point estimates of Equation 1 in Panel 1a, and of Equation 2 in Panel 1b. The share of males is discretized and juries with 50 percent males are used as omitted category. Standard errors are clustered at the committee level.

Appendix Figure 2: Standardized Differences with Respect to Gender Balanced Jury by Candidates' Gender and Jury Member Fixed Effect

(a) Females

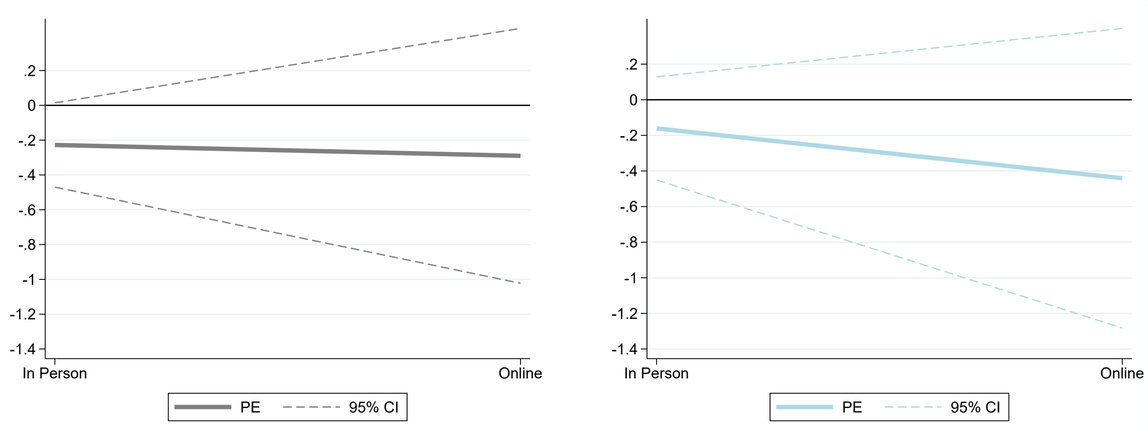
(b) Males



Notes. These figures present the point estimates of Equation 2 using the share of males discretized. Juries with 50 percent males are used as omitted category. Panel 2a conditions on female students, whereas Panel 2b on male students. Standard errors are clustered at the committee level.

Appendix Figure 3: Heterogeneous Effects by Taking Exam Online

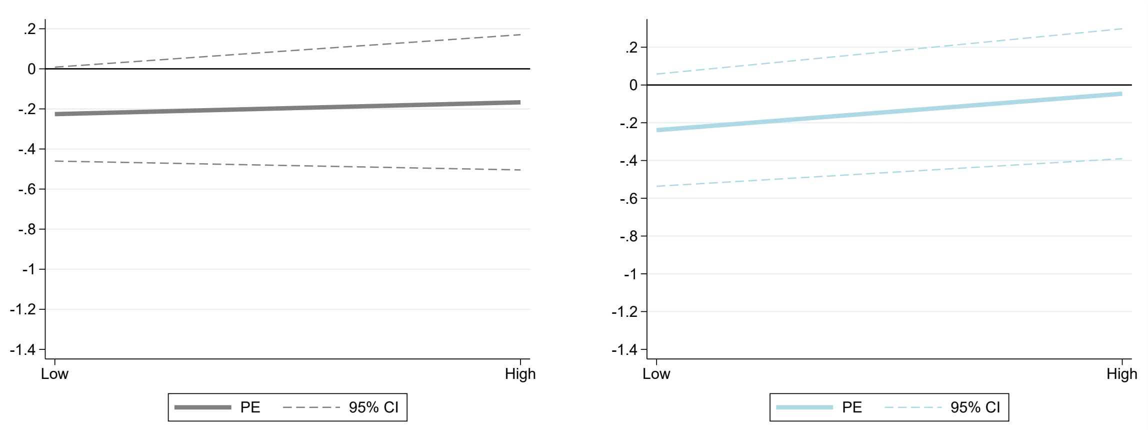
(a) Male Candidates (b) Female Candidates



Notes. These figures present the point estimates of Equation 2 including an interaction term between the share of males and the dummy variable for taking the oral core exam online.The plotted point estimates correspond to the coefficient attached to the share of males (when the x-axis takes the value of zero), and that same coefficient plus the coefficient attached to the interaction (when the x-axis takes the value of one). All specifications include the jury member fixed effects and the full set of controls, equivalent to column (6) of Table 1 Standard errors are clustered at the committee level. 95% confidence intervals are plotted.

Appendix Figure 4: Heterogeneous Effects by Oral Foreign Language Test Score

(a) Male Candidates (b) Female Candidates



Notes. These figures present the point estimates of Equation 2 including an interaction term between the share of males and the test score in the oral foreign language component. The plotted point estimates correspond to the coefficient attached to the share of males (when the x-axis takes the value of zero), and that same coefficient plus the coefficient attached to the interaction (when the x-axis takes the value of one). All specifications include the jury member fixed effects and the full set of controls, equivalent to column (6) of Table 1 Standard errors are clustered at the committee level. 95% confidence intervals are plotted.

Appendix Table 1: Descriptive Statistics

Variable	Observations (1)	Mean (2)	Median (3)	Std. Deviation (4)	Min. (5)	Max. (6)
A) Committees						
Share of Males (β)	2,611	0.49	0.50	0.25	0.00	1.00
1(Morning Sesion)	2,611	0.50	0.00	0.50	0.00	1.00
1(Online)	2,611	0.08	0.00	0.28	0.00	1.00
1(Site 1)	2,611	0.37	0.00	0.48	0.00	1.00
1(Site 2)	2,611	0.55	1.00	0.50	0.00	1.00
1(3 Juries)	2,611	0.14	0.00	0.35	0.00	1.00
Share of In-House Juries	2,611	0.61	0.50	0.25	0.00	1.00
Share of Faculty Members	2,611	0.26	0.00	0.29	0.00	1.00
B) Students						
1(Female)	2,611	0.39	0.00	0.49	0.00	1.00
Literacy Written Exam (sd)	2,611	-0.00	0.02	1.00	-3.61	2.35
Math Written Exam (sd)	2,611	-0.00	0.05	1.00	-3.99	3.19
English Written Exam (sd)	2,611	-0.00	-0.02	1.00	-3.74	2.66
English Oral Exam (sd)	2,610	-0.00	0.06	1.00	-3.82	1.72
Admission Exam (sd)	2,611	-0.00	0.25	1.00	-3.47	1.50
C) Jury Members						
1(Jury Male)	694	0.47	0.00	0.50	0.00	1.00
1(Jury Faculty)	694	0.16	0.00	0.37	0.00	1.00
1(Jury Student)	694	0.05	0.00	0.23	0.00	1.00
1(Jury in House)	694	0.42	0.00	0.49	0.00	1.00
Number of Students	694	8.05	5.00	6.39	1.00	42.00
Number of Committees	694	2.65	2.00	2.22	1.00	15.00

Appendix Table 2: Description of Committees

Variable	Obs. (1)	Mean (2)	Median (3)	Std. Dev. (4)	Min. (5)	Max. (6)	Average By Share of Males (%) 0.00 (7)	0.33 (8)	0.50 (9)	0.66 (10)	1.00 (11)
Number of Males	862	1.03	1.00	0.57	0.00	3.00	0.00	1.00	1.00	2.00	2.02
1(Jury Male)	862	0.48	0.50	0.26	0.00	1.00	0.00	0.33	0.50	0.67	1.00
Number of Juries	862	2.14	2.00	0.35	2.00	3.00	2.01	3.00	2.00	3.00	2.04
Share of In-House Juries	862	0.61	0.50	0.26	0.00	1.00	0.59	0.58	0.60	0.59	0.70
Share of Faculty Members	862	0.26	0.00	0.29	0.00	1.00	0.17	0.26	0.28	0.20	0.30
1(3 Juries)	862	0.14	0.00	0.35	0.00	1.00	0.01	1.00	0.00	1.00	0.04
Admission Exam (sd)	862	-0.01	0.01	0.64	-2.73	1.50	0.04	-0.05	0.01	-0.04	-0.15
Total (N)							123	56	534	59	90

Appendix Table 3: Balance on Individual Characteristics

	1(Female)		1(Morning)		1(Online)		1(Site 1)		1(Site 2)		1(3 Juries)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Share of Males	0.017	-0.031	-0.062	-0.005	0.048	-0.019	0.018	-0.050	-0.066	0.069**	0.077*	0.037
	(0.036)	(0.046)	(0.072)	(0.089)	(0.041)	(0.049)	(0.069)	(0.039)	(0.072)	(0.034)	(0.042)	(0.061)
Observations	2,611	5,584	2,611	5,584	2,611	5,584	2,611	5,584	2,611	5,584	2,611	5,584
Jury FE		Yes		Yes		Yes		Yes		Yes		Yes

Note: This table presents the results of a regression of the share of male jury members on the outcome described at the top of the table. Odd columns does not include any covariate and are estimated at the student level. Even columns are estimated at the student-jury member level and include jury member fixed effects. Standard errors are clustered at the committee level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Appendix Table 4: Balance on Test Scores

	Written Exams						Oral Exam:	
	Literacy		Numeracy		Foreign Lang.		Foreign Lang.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share of Males	0.108	0.005	0.003	-0.035	0.125*	0.070	0.099	0.076
	(0.076)	(0.087)	(0.073)	(0.092)	(0.074)	(0.094)	(0.083)	(0.102)
Observations	2,611	5,584	2,611	5,584	2,611	5,584	2,610	5,582
Jury FE		Yes		Yes		Yes		Yes

Note: This table presents the results of a regression of the share of male jury members on the outcome described at the top of the table. Odd columns does not include any covariate and are estimated at the student level. Even columns are estimated at the student-jury member level and include jury member fixed effects. Standard errors are clustered at the committee level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Appendix Table 5: Effects of Jury Gender Composition on Oral Test Scores Only two Jury-Members

	(1)	(2)	(3)	(4)	(5)	(6)
A) No Jury Member Fixed-Effect						
Share of Males (β)	-0.151*	-0.201**	-0.135*	-0.165	-0.148*	-0.182*
	(0.081)	(0.086)	(0.079)	(0.103)	(0.078)	(0.102)
Share of Males*1(female) (γ)		0.136*		0.076		0.096
		(0.076)		(0.166)		(0.163)
Women ($\beta + \gamma$)		-0.0656		-0.0883		-0.0865
p-value		0.485		0.491		0.493
Observations	2,248	2,248	2,248	2,248	2,247	2,247
B) Jury Member Fixed-Effect						
Share of Males (β)	-0.215**	-0.257***	-0.218***	-0.254**	-0.239***	-0.270**
	(0.084)	(0.089)	(0.083)	(0.110)	(0.084)	(0.111)
Share of Males*1(female) (γ)		0.125		0.089		0.091
		(0.087)		(0.185)		(0.180)
Women ($\beta + \gamma$)		-0.132		-0.165		-0.180
p-value		0.195		0.243		0.190
Observations	4,495	4,495	4,495	4,495	4,493	4,493
Candidate Controls			Yes	Yes	Yes	Yes
Baseline Test Scores					Yes	Yes

Note: This table presents the results of the estimation of Equation 1 (Panel A) and Equation 2 (Panel B) using the standardized oral exam test score as dependent variable and excluding students assigned to more than two members. The share of males corresponds to the percentage of male jury members in each committee. Columns (1) and (2) have no controls, columns (3) and (4) include candidate-level controls, and columns (5) and (6) include candidate-level controls and test score measures. Candidate-level controls include gender, if the exam was performed in the morning, site fixed effects (online, site one, and site two), if the committee had three members, the share of jury members that are part of the school, and the share of jury members that are part of the school's faculty. Test score measures include the candidate's standardized result in the blind (reading, math, and English) and non-blind exams (oral English exam). Specifications in even columns fully interact the share of males and the controls with a dummy variable of whether or not the candidate is female. Standard errors are clustered at the committee level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Appendix Table 6: Effects of Jury Gender Composition on Oral Test Scores using Sample of Candidates for Admission into Further Years

	(1)	(2)	(3)	(4)
Share of Males	-0.254***	-0.253***	-0.250**	-0.208*
	(0.092)	(0.092)	(0.110)	(0.110)
Observations	1,342	1,342	2,876	2,876
Candidate Controls		Yes		Yes
Jury FE			Yes	Yes

Note: Estimations performed in the sample of students who took the oral admission exam in second and third years. The outcome corresponds to the standardized oral exam test score. The share of males corresponds to the percentage of male jury members in each committee. Candidate-level controls include gender, if the exam was performed in the morning, site fixed effects (online, site one, and site two), if the committee had three members, the share of jury members that are part of the school, and the share of jury members that are part of the school's faculty. Standard errors are clustered at the committee level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Appendix Table 7: Effects of Jury Gender Composition on Difference Between Blind and Non-Blind Scores

	(1)	(2)	(3)	(4)	(5)	(6)
A) No Jury Member Fixed-Effect						
Share of Males (β)	-0.180**	-0.253***	-0.166**	-0.223**	-0.160**	-0.208**
	(0.077)	(0.082)	(0.075)	(0.098)	(0.075)	(0.098)
Share of Males*1(female) (γ)		0.193***		0.150		0.130
		(0.071)		(0.157)		(0.155)
Women ($\beta + \gamma$)		-0.0608		-0.0730		-0.0779
p-value		0.490		0.545		0.516
Observations	2,611	2,611	2,611	2,611	2,610	2,610
B) Jury Member Fixed-Effect						
Share of Males (β)	-0.210***	-0.291***	-0.217***	-0.298***	-0.219***	-0.289***
	(0.081)	(0.087)	(0.079)	(0.107)	(0.080)	(0.108)
Share of Males*1(female) (γ)		0.230***		0.202		0.180
		(0.081)		(0.173)		(0.169)
Women ($\beta + \gamma$)		-0.0607		-0.0964		-0.110
p-value		0.516		0.448		0.383
Observations	5,584	5,584	5,584	5,584	5,582	5,582
Candidate Controls			Yes	Yes	Yes	Yes
Baseline Test Scores					Yes	Yes

Note: This table presents the results of the estimation of Equation 1 (Panel A) and Equation 2 (Panel B) using the differences between the blind and non-blind test scores as dependent variable. The share of males corresponds to the percentage of male jury members in each committee. Columns (1) and (2) have no controls, columns (3) and (4) include candidate-level controls, and columns (5) and (6) include candidate-level controls and test score measures. Candidate-level controls include gender, if the exam was performed in the morning, site fixed effects (online, site one, and site two), if the committee had three members, the share of jury members that are part of the school, and the share of jury members that are part of the school's faculty. Test score measures include the candidate's standardized result in the blind (reading, math, and English) and non-blind exams (oral English exam). Specifications in even columns fully interact the share of males and the controls with a dummy variable of whether or not the candidate is female. Standard errors are clustered at the committee level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Appendix Table 8: Heterogeneous Effects by Baseline Characteristics and Gender of the Candidates

	Online (1)	Oral Foreign Lang. (2)	Literacy (3)	Numeracy (4)	Foreign Lang. (5)
A) Overall					
Interaction	-0.110	0.137	0.057	-0.161**	0.070
	(0.246)	(0.085)	(0.083)	(0.079)	(0.083)
Share of Males (β)	-0.206**	-0.220***	-0.217***	-0.211***	-0.216***
	(0.084)	(0.080)	(0.080)	(0.079)	(0.080)
Observations	5,582	5,582	5,582	5,582	5,582
B) Male Candidates					
Interaction	-0.062	0.059	0.191	-0.068	0.059
	(0.386)	(0.115)	(0.117)	(0.109)	(0.116)
Share of Males (β)	-0.228*	-0.226*	-0.210*	-0.217*	-0.223*
	(0.123)	(0.119)	(0.120)	(0.123)	(0.120)
Observations	3,340	3,340	3,340	3,340	3,340
C) Female Candidates					
Interaction	-0.280	0.193	-0.233	-0.280*	0.110
	(0.446)	(0.158)	(0.144)	(0.150)	(0.165)
Share of Males (β)	-0.161	-0.239	-0.150	-0.275*	-0.217
	(0.148)	(0.151)	(0.146)	(0.155)	(0.145)
Observations	2,050	2,050	2,050	2,050	2,050
Candidate Controls	Yes	Yes	Yes	Yes	Yes
Baseline Test Scores	Yes	Yes	Yes	Yes	Yes

Note: These figures present the results of estimating Equation 2 including an interaction term between the share of males and the baseline characteristic described in the title of column. All specifications include candidate-level controls and test score measures. Candidate-level controls include gender, if the exam was performed in the morning, site fixed effects (online, site one, and site two), if the committee had three members, the share of jury members that are part of the school, and the share of jury members that are part of the school's faculty. Test score measures include the candidate's standardized result in the blind (reading, math, and English) and non-blind exams (oral English exam).*** p<0.01, ** p<0.05, * p<0.1..